Appendix for "The Binomial Block Bootstrap Estimator for Evaluating Loss on Dependent Clusters"

A Proofs

A.1 Lemma 3.1

Proof. Let A_j denote column j of matrix A. The entries of A_j correspond to polynomial $g_j(q) = \binom{n'}{j} q^j (1-q)^{(n'-j)}$ evaluated at points $q = p_0, p_1, \ldots, p_m$. First, we show polynomials $g(q) = \{g_0(q), \ldots, g_{n'}(q)\}$ are linearly independent.

We look for a non-trivial solution $\kappa \in \mathbb{R}^{(n'+1)}$ to $\langle \kappa, g(q) \rangle = 0, \forall q \in [0,1]$. The binomial coefficient is a constant in each polynomial and can be dropped. Expanding and collecting terms,

$$\begin{aligned} 0 &= \kappa_0 (1-q)^{n'} + \kappa_1 q^1 (1-q)^{n'-1} + \ldots + \kappa_{n'} q^{n'} \\ &= \sum_{i=0}^{n'} q^i \sum_{j=0}^{i} \kappa_j \binom{n'-j}{i-j} (-1)^j \end{aligned}$$

for all $q \in \left[0, \frac{n-1}{n}\right]$ which implies

$$0 = \sum_{j=0}^{i} \kappa_j \binom{n'-j}{i-j} (-1)^j$$

for all $i \in \{0, 1, \ldots, n'\}$ and all $q \in \left[0, \frac{n-1}{n}\right]$. Clearly, $\kappa_0 = 0$, and the remainder of the terms follow by induction to $\kappa = \vec{0}$. Thus the polynomials $\{g_0, \ldots, g_{n'}\}$ are linearly independent.

Next, the polynomials $\{g_0, \ldots, g_{n'}\}$ are unisolvent by the unisolvence theorem, which implies the vectors

$$\begin{bmatrix} g_0(p_0) \\ g_0(p_1) \\ \vdots \\ g_0(p_m) \end{bmatrix}, \begin{bmatrix} g_1(p_0) \\ g_1(p_1) \\ \vdots \\ g_1(p_m) \end{bmatrix}, \ldots, \begin{bmatrix} g_{n'}(p_0) \\ g_{n'}(p_1) \\ \vdots \\ g_{n'}(p_m) \end{bmatrix}, \tag{7}$$

are also linearly independent for any unique $p_0, \ldots, p_m$, $m \geq n'$. Thus, matrix A is full rank. □

A.2 Theorem 3.2

A.2.1 Statement 1

Proof. Without loss of generality, we prove the corruption direction from $\mathcal{V}$ to $\mathcal{T}$. The empirical loss of KNOWNUNIDIRECTIONAL at corruption level p_i is $\bar{b}_i$.

$$\begin{aligned}
\bar{b}_i &= \frac{1}{t|\hat{\mathcal{V}}|} \sum_{\mathcal{T}' \in \{\mathcal{T}'_0, \ldots, \mathcal{T}'_t\}} \sum_{(x,y) \in \hat{\mathcal{V}}} \ell(y, f(x \mid \mathcal{T}')) \\
&= \frac{1}{t|\hat{\mathcal{V}}|} \sum_{(x,y) \in \hat{\mathcal{V}}} \sum_{j=0}^{n'} \sum_{\mathcal{T}' : |\mathcal{T}' \cap \mathcal{V}| = j, \mathcal{T}' \in \{\mathcal{T}'_0, \ldots, \mathcal{T}'_t\}} \ell(y, f(x \mid \mathcal{T}')) \\
&\xrightarrow[t \to \infty]{p} \frac{1}{|\hat{\mathcal{V}}|} \sum_{(x,y) \in \hat{\mathcal{V}}} \sum_{j=0}^{n'} A_{ij} \mathbb{E}_{\mathcal{T}''(j)} \ell(y, f(x \mid \mathcal{T}'')) \\
&\xrightarrow[|\mathcal{V}|, |\mathcal{T}| \to \infty]{p} \mathop{\mathbb{E}}_{(x,y) \sim P_{\mathcal{V}}} \sum_{j=0}^{n'} A_{ij} \mathbb{E}_{\mathcal{T}'''(j)} \ell(y, f(x \mid \mathcal{T}''')) \\
&= \sum_{j=0}^{n'} A_{ij} e_j = b_i
\end{aligned}$$

where A_{ij} is the probability of sampling j samples from $\mathcal{V}$ at corruption p_i as defined in Eq. 5 and

$$\begin{aligned}
\mathcal{T}''(j) &= \left\{ \{\mathcal{T}'''' \overset{n'-j}{\sim} \mathcal{T}\} \cup \{\mathcal{V}'' \overset{j}{\sim} \mathcal{V}\} \right\} \\
\mathcal{T}'''(j) &= \left\{ \{\mathcal{T}'''' \overset{n'-j}{\sim} P_{\mathcal{T}}\} \cup \{\mathcal{V}'' \overset{j}{\sim} P_{\mathcal{V}}\} \right\} \\
e_j &= \mathbb{E}_{\mathcal{T}'''(j), (x,y) \sim P_{\mathcal{V}}} \ell(y, f(x \mid \mathcal{T}'))
\end{aligned}$$

The proof can be understood as splitting the t bootstraps into bins $j = 0, \ldots, n'$ each with probability A_{ij}. Then $\hat{e} = A(A^\intercal A)^{-1} A^\intercal \bar{b} \xrightarrow{p} A(A^\intercal A)^{-1} A^\intercal b = e$ by the continuous mapping theorem and Lemma 3.1. We require $|\mathcal{V}|, |\mathcal{T}| \to \infty$ sufficiently faster than $t \to \infty$, such that the probability of resampling the same data sample in Algorithm 1 also goes to 0. □

A.2.2 Statement 2

Proof. Recall $\hat{e} = A(A^\intercal A)^{-1} A^\intercal b$ and $\hat{e} = A(A^\intercal A)^{-1} A^\intercal \bar{b}$. Then

$$\begin{aligned}
\mathbb{E}[\hat{e}] &= \mathbb{E}[A(A^\intercal A)^{-1} A^\intercal \bar{b}] \\
&= A(A^\intercal A)^{-1} A^\intercal \mathbb{E}[\bar{b}] \\
&= A(A^\intercal A)^{-1} A^\intercal b \\
&= e
\end{aligned}$$

for finite t but infinitely large $|\hat{\mathcal{V}}|$ and $|\hat{\mathcal{T}}|$. For finite sample sets, bias may be introduced due to resampling the same data sample more than once. □

A.2.3 Statement 3

Proof. We consider the case where A is square for analysis simplicity. Let $\sigma^2_{\hat{b}_i}$ be the variance of $\hat{b}_i$ in line 13 of Algorithm 1. Then

$$\begin{aligned}
\text{Var}(\hat{e}_0) &= \sum_{i=0}^{n'} (A^{-1})^2_{0i} \text{Var}(\hat{b}_i) \\
&= \sum_{i=0}^{n'} (A^{-1})^2_{0i} \frac{\sigma^2_{\hat{b}_i}}{t} \\
&= \frac{1}{t} u^\intercal (A^{-1})^\intercal \Sigma A^{-1} u
\end{aligned}$$

where Σ is the diagonal matrix with entires $\Sigma_{ii} = \sigma^2_{\hat{b}_i}$ and u denotes a standard unit vector, such that $u_0 = 1$. Let $z = A^{-1}u$. Now solving for z

$$Az = u$$

$$\binom{n'}{j} \sum_{i=0}^{n'} p_i^j (1-p_i)^{n'-j} e_i = \begin{cases} 1 & j = 0 \\ 0 & j = 1, \ldots, n' \end{cases}$$

For $j = n'$, note $\sum_{i=0}^{n'} p_i^{n'} z_i = 0$. Then expanding terms for $j = n' - 1$

$$\sum_{i=0}^{n'} p_i^{n'-1}(1-p_i) z_i = 0$$

$$\sum_{i=0}^{n'} p_i^{n'-1} z_i - p_i^{n'} z_i = 0$$

$$\sum_{i=0}^{n'} p_i^{n'-1} z_i = 0$$

Continuing these expansion and substitution steps for $j = n' - 1, n' - 2, \ldots, 1, 0$ results in

$$\sum_{i=0}^{n'} p_i^k z_i = u_k \quad \text{for } k = 0, \ldots, n'$$

which is a transposed Vandermonde system

$$\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ p_0 & p_1 & p_2 & \cdots & p_{n'} \\ p_0^2 & p_1^2 & p_2^2 & \cdots & p_{n'}^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_0^{n'} & p_1^{n'} & p_2^{n'} & \cdots & p_{n'}^{n'} \end{bmatrix} \begin{bmatrix} z_0 \\ z_1 \\ z_2 \\ \vdots \\ z_{n'} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$W^\intercal z = u$$

$$z = (W^{-1})^\intercal u$$

Thus, $\text{Var}(\hat{e}_0) = \frac{1}{t} u^\intercal W^{-1} \Sigma (W^{-1})^\intercal u$ and the inverse of the Vandermonde matrix is known [30]

$$(W^{-1})_{ij} = \begin{cases} (-1)^i \left(\frac{\sum_{0 \le m_0 < \cdots < m_{n'-1} \le n', m_0, \ldots, m_{n'-i} \neq j} p_{m_0} \cdots p_{m_{n'-i}}}{\prod_{0 \le m \le n', m \neq j} (p_m - p_j)} \right) & \text{for } 0 \le i \le n' \\ \frac{1}{\prod_{0 \le m \le n', m \neq j} (p_m - p_j)} & \text{for } k = n' \end{cases}$$

which gives the result. □

A.3 Lemma 3.5

Proof. Let $A_{k+n'_{\mathcal{T}}l}$ denote column $k + n'_{\mathcal{T}}l$ of matrix A. The entries of $A_{k+n'_{\mathcal{T}}l}$ correspond to polynomial $g_{k+n'_{\mathcal{T}}l}(q_1, q_2) = \binom{n'_{\mathcal{T}}}{k} q_1^k (1-q_1)^{(n'_{\mathcal{T}}-k)} \binom{n'_{\mathcal{V}}}{l} q_2^l (1-q_2)^{(n'_{\mathcal{V}}-l)}$ evaluated at points $q_1 = p_{\mathcal{T},0}, p_{\mathcal{T},1}, \ldots, p_{\mathcal{T},m}$ and $q_2 = p_{\mathcal{V},0}, p_{\mathcal{V},1}, \ldots, p_{\mathcal{V},n}$. First, we show polynomials $g = \{g_0, \ldots, g_{(n'_{\mathcal{T}}+1)(n'_{\mathcal{V}}+1)-1}\}$ are linearly independent.

We look for a non-trivial solution $K \in \mathbb{R}^{(n'_{\mathcal{T}}+1, n'_{\mathcal{V}}+1)}$ to $\langle c, g(q_1, q_2)\rangle = 0, \forall q_1, q_2 \in [0,1]$ where κ is a flattened version of K. The binomial coefficient is a constant in each polynomial and can be dropped. Expanding and collecting terms,

$$
\begin{aligned}
0 &= \sum_{i=0}^{n'_{\mathcal{T}}} \sum_{j=0}^{n'_{\mathcal{V}}} K_{ij} q_1^i (1-q_1)^{n'_{\mathcal{T}}-i} q_2^j (1-q_2)^{n'_{\mathcal{V}}-j} \\
&= \sum_{j=0}^{n'_{\mathcal{V}}} q_2^j (1-q_2)^{n'_{\mathcal{V}}-j} \sum_{i=0}^{n'_{\mathcal{T}}} q_1^i \sum_{k=0}^{i} K_{kj} \binom{n'_{\mathcal{T}}-k}{i-k} (-1)^k \\
&= \sum_{i=0}^{n'_{\mathcal{T}}} q_1^i \sum_{k=0}^{i} \binom{n'_{\mathcal{T}}-k}{i-k} (-1)^k \sum_{j=0}^{n'_{\mathcal{V}}} K_{kj} q_2^j (1-q_2)^{n'_{\mathcal{V}}-j} \\
&= \sum_{i=0}^{n'_{\mathcal{T}}} q_1^i \sum_{k=0}^{i} \binom{n'_{\mathcal{T}}-k}{i-k} \sum_{j=0}^{n'_{\mathcal{V}}} q_2^j \sum_{l=0}^{j} K_{kl} \binom{n'_{\mathcal{V}}-l}{j-l} (-1)^{k+l}
\end{aligned}
$$

which implies,

$$
0 = \sum_{k=0}^{i} \sum_{l=0}^{j} K_{kl} \binom{n'_{\mathcal{T}}-k}{i-k} \binom{n'_{\mathcal{V}}-l}{j-l} (-1)^{k+l}
$$

for all $i \in \{0, 1, \ldots, n'_{\mathcal{T}}\}$ and $j \in \{0, 1, \ldots, n'_{\mathcal{V}}\}$. When $i = 0, j = 0 \Rightarrow K_{00} = 0$ and the remainder of the terms follow by induction to $K = \mathbb{0}^{(n'_{\mathcal{T}}+1)\times(n'_{\mathcal{T}}+1)}$. Thus the polynomials $\{g_0, \ldots, g_{(n'_{\mathcal{T}}+1)(n'_{\mathcal{V}}+1)-1}\}$ are linearly independent and A is full rank by the unisolvence theorem. □

A.4 Theorem 3.6

Proof. The empirical loss at corruption levels $(p_{\mathcal{T},i}, p_{\mathcal{V},j})$ is $\bar{b}_{ij}$ in Algorithm 3.

$$
\begin{aligned}
\bar{b}_{ij} &= \frac{1}{t n'_{\mathcal{V}}} \sum_{\mathcal{T}' \in \{\mathcal{T}'_0, \ldots, \mathcal{T}'_t\}} \sum_{\mathcal{V}' \in \{\mathcal{V}'_0, \ldots, \mathcal{V}'_t\}} \sum_{(x,y) \in \mathcal{V}'} \ell(y, f(x \mid \mathcal{T}')) \\
&= \frac{1}{t n'_{\mathcal{V}}} \sum_{l=0}^{n'_{\mathcal{V}}} \sum_{\mathcal{V}' : |\mathcal{V}' \cap \mathcal{T}| = l, \mathcal{V}' \in \{\mathcal{V}'_0, \ldots, \mathcal{V}'_t\}} \sum_{(x,y) \in \mathcal{V}'} \\
&\qquad \sum_{k=0}^{n'_{\mathcal{T}}} \sum_{\mathcal{T}' : |\mathcal{T}' \cap \mathcal{V}| = k, \mathcal{T}' \in \{\mathcal{T}'_0, \ldots, \mathcal{T}'_t\}} \ell(y, f(x \mid \mathcal{T}')) \\
&\xrightarrow[t \to \infty]{p} \sum_{l=0}^{n'_{\mathcal{V}}} \sum_{k=0}^{n'_{\mathcal{T}}} A_{ijkl} \mathbb{E}_{\mathcal{T}''(k), (x,y) \in \mathcal{V}''(l)} \ell(y, f(x \mid \mathcal{T}'')) \\
&\xrightarrow[|\mathcal{V}|, |\mathcal{T}| \to \infty]{p} \sum_{l=0}^{n'_{\mathcal{V}}} \sum_{k=0}^{n'_{\mathcal{T}}} A_{ijkl} \mathbb{E}_{\mathcal{T}'''(k), (x,y) \in \mathcal{V}'''(l)} \ell(y, f(x \mid \mathcal{T}''')) \\
&= \sum_{l=0}^{n'_{\mathcal{V}}} \sum_{k=0}^{n'_{\mathcal{T}}} A_{ijkl} x_{kl}
\end{aligned} \tag{8}
$$

where A_{ijkl} is the probability of k corrupted samples in $\mathcal{T}'$ and l corrupted samples in $\mathcal{V}'$ at corruption levels $p_{\mathcal{T},i}$ and $p_{\mathcal{V},j}$, i.e.

$$A_{ijkl} = \mathbb{P}(\text{Binomial}(n'_{\mathcal{T}}, p_{\mathcal{T},i}) = k)\mathbb{P}(\text{Binomial}(n'_{\mathcal{V}}, p_{\mathcal{V},j}) = l)$$

and

$$\begin{aligned}
\mathcal{T}''(k) &= \left\{ \{\mathcal{T}'''' \overset{n'_{\mathcal{T}}-k}{\sim} \mathcal{T}\} \cup \{\mathcal{V}'''' \overset{k}{\sim} \mathcal{V}\} \right\} \\
\mathcal{V}''(l) &= \left\{ \{\mathcal{T}'''' \overset{l}{\sim} \mathcal{T}\} \cup \{\mathcal{V}'''' \overset{n'_{\mathcal{V}}-l}{\sim} \mathcal{V}\} \right\} \\
\mathcal{T}'''(k) &= \left\{ \{\mathcal{T}'''' \overset{n'_{\mathcal{T}}-k}{\sim} P_{\mathcal{T}}\} \cup \{\mathcal{V}'''' \overset{k}{\sim} P_{\mathcal{V}}\} \right\} \\
\mathcal{V}'''(l) &= \left\{ \{\mathcal{T}'''' \overset{l}{\sim} P_{\mathcal{T}}\} \cup \{\mathcal{V}'''' \overset{n'_{\mathcal{V}}-l}{\sim} P_{\mathcal{V}}\} \right\} \\
x_{kl} &= \mathbb{E}_{\mathcal{T}'''(k),(x,y)\in\mathcal{V}'''(l)} \ell(y, f(x \mid \mathcal{T}'''))
\end{aligned}$$

We flatten x, b and appropriately reshape the tensor A into a matrix such that Eq. 8 is always satisfied in the linear system $Ax = b$. Then, $\hat{e} = A(A^\intercal A)^{-1}A^\intercal \bar{b} \xrightarrow{p} A(A^\intercal A)^{-1}A^\intercal b = x$ by the continuous mapping theorem and Lemma 3.1. □

A.5 Theorem 3.7

Proof. We begin by proving the convergence of $p^*_{\mathcal{T},0}$ and $p^*_{\mathcal{V},0}$. Let $n_{\mathcal{T}} = |\hat{\mathcal{T}}|$ and $n_{\mathcal{V}} = |\hat{\mathcal{V}}|$. In Algorithm 4, $(p^*_{\mathcal{T},0}, p^*_{\mathcal{V},0}) = \arg\min_{p_{\mathcal{T},0}\in\left\{\frac{0}{n_{\mathcal{T}}}, \frac{1}{n_{\mathcal{T}}}, \ldots, \frac{n_{\mathcal{T}}-1}{n_{\mathcal{T}}}\right\}, p_{\mathcal{V},0}\in\left\{\frac{0}{n_{\mathcal{V}}}, \frac{1}{n_{\mathcal{V}}}, \ldots, \frac{n_{\mathcal{V}}-1}{n_{\mathcal{V}}}\right\}} g^{(t)}(p_{\mathcal{T},0}, p_{\mathcal{V},0})$, where the function $g^{(i)}$ is defined as

$$g^{(i)}(p_{\mathcal{T},0}, p_{\mathcal{V},0}) = \begin{cases} ||A(p_{\mathcal{T},0}, p_{\mathcal{V},0})(A^\intercal(p_{\mathcal{T},0}, p_{\mathcal{V},0})A(p_{\mathcal{T},0}, p_{\mathcal{V},0}))^{-1}A^\intercal(p_{\mathcal{T},0}, p_{\mathcal{V},0})\bar{b}^{(i)} - \bar{b}^{(i)}||_2^2 \\ \qquad\qquad \text{if } p_{\mathcal{T},0} \in \left[0, \frac{n_{\mathcal{T}}-1}{n_{\mathcal{T}}}\right], p_{\mathcal{V},0} \in \left[0, \frac{n_{\mathcal{V}}-1}{n_{\mathcal{V}}}\right] \\ \infty \qquad\qquad \text{else} \end{cases}$$

$$g(p_{\mathcal{T},0}, p_{\mathcal{V},0}) = \begin{cases} ||A(p_{\mathcal{T},0}, p_{\mathcal{V},0})(A^\intercal(p_{\mathcal{T},0}, p_{\mathcal{V},0})A(p_{\mathcal{T},0}, p_{\mathcal{V},0}))^{-1}A^\intercal(p_{\mathcal{T},0}, p_{\mathcal{V},0})b - b||_2^2 \\ \qquad\qquad \text{if } p_{\mathcal{T},0} \in \left[0, \frac{n_{\mathcal{T}}-1}{n_{\mathcal{T}}}\right], p_{\mathcal{V},0} \in \left[0, \frac{n_{\mathcal{V}}-1}{n_{\mathcal{V}}}\right] \\ \infty \qquad\qquad \text{else} \end{cases}$$

Both g and the sequence of functions $\{g^{(0)}, g^{(1)}, \dots\}$ are level-bounded, lower semi-continuous and proper. By Lemma A.1, $g^{(i)} \xrightarrow{e} g$ where $\xrightarrow{e}$ denotes convergence in epigraph. Thus, $residual = \min_{p_{\mathcal{T},0}\in\left[0, \frac{n_{\mathcal{T}}-1}{n_{\mathcal{T}}}\right], p_{\mathcal{V},0}\in\left[0, \frac{n_{\mathcal{V}}-1}{n_{\mathcal{V}}}\right]} g^{(t)}(p_{\mathcal{T},0}, p_{\mathcal{V},0}) \xrightarrow{p} \min_{p_{\mathcal{T},0}, p_{\mathcal{V},0}} g(p_{\mathcal{T},0}, p_{\mathcal{V},0})$ [28]. We know at least one perfect solution $g(p_{\mathcal{T},0}, p_{\mathcal{V},0}) = 0$ exists, that this solution is unique (by Assumption 2) and that this solution is in $\left\{0, \frac{1}{n_{\mathcal{T}}}, \ldots, \frac{n_{\mathcal{T}}-1}{n_{\mathcal{T}}}\right\} \mathcal{X} \left\{0, \frac{1}{n_{\mathcal{V}}}, \ldots, \frac{n_{\mathcal{V}}-1}{n_{\mathcal{V}}}\right\}$. Thus, $p^*_{\mathcal{T},0} \xrightarrow{p} p_{\mathcal{T},0}$, $p^*_{\mathcal{V},0} \xrightarrow{p} p_{\mathcal{V},0}$ and $residual \xrightarrow{p} 0$. □

Assumption 2. *b is independent of the columns of $A(p_{\hat{\mathcal{T}},0}, p_{\hat{\mathcal{V}},0})$ (except, obviously, at $p_{\hat{\mathcal{T}},0}, p_{\hat{\mathcal{V}},0} = p_{\mathcal{T},0}, p_{\mathcal{V},0}$). This is a very weak assumption when choosing $m >> n_{\mathcal{T}}n_{\mathcal{V}}$. It is unlikely the loss vector b happens to fall in the column space of A.*

Lemma A.1. *$g^{(i)} \xrightarrow{e} g$, where we use $\xrightarrow{e}$ to denote convergence in epigraph.*

Proof. Let $x = (p_{\mathcal{T},0}, p_{\mathcal{V},0})$. Then the proof follows exactly from the proof of Lemma 3.4. □

B Algorithms

Algorithm 3: B3 — Bidirectional leakage with known probabilities

procedure KNOWNBIDIRECTIONAL($f, \hat{\mathcal{T}}, \hat{\mathcal{V}}, p_{\mathcal{T},0}, p_{\mathcal{V},0}, n'_{\mathcal{T}}, n'_{\mathcal{V}}, t$)
  $\bar{b} \leftarrow \mathbb{0}^{(n_{\mathcal{T}'}+1)\times(n_{\mathcal{V}'}+1)}$
  for p_i in $\{p_{\mathcal{T},0}, p_{\mathcal{T},0}+\delta_{\mathcal{T}}, p_{\mathcal{T},0}+2\delta_{\mathcal{T}}, \ldots, 1\}$ **do** ▷ Choose $\delta_{\mathcal{T}} > 0$ s.t. $|\{p_i\}| > n_{\mathcal{T}}$
    $p'_{\mathcal{T}} \leftarrow \frac{p_i+p_{\mathcal{V},0}-1}{p_{\mathcal{T},0}+p_{\mathcal{V},0}-1}$
    for p_j in $\{p_{\mathcal{V},0}, p_{\mathcal{V},0}+\delta_{\mathcal{V}}, p_{\mathcal{V},0}+2\delta_{\mathcal{V}}, \ldots, 1\}$ **do** ▷ Choose $\delta_{\mathcal{V}} > 0$ s.t. $|\{p_j\}| > n_{\mathcal{V}}$
      $p'_{\mathcal{V}} \leftarrow \frac{p_j+p_{\mathcal{T},0}-1}{p_{\mathcal{V},0}+p_{\mathcal{T},0}-1}$
      for $k \leftarrow 1$ to t **do**
        $\mathcal{T}'_k \overset{n'_{\mathcal{T}}}{\sim} M_{\hat{\mathcal{T}},\hat{\mathcal{V}}}(p'_{\mathcal{T}}, 1-p'_{\mathcal{T}})$ ▷ M is a mixture distribution[2]
        $\mathcal{V}'_k \overset{n'_{\mathcal{V}}}{\sim} M_{\hat{\mathcal{V}},\hat{\mathcal{T}}}(p'_{\mathcal{V}}, 1-p'_{\mathcal{V}})$
        $\hat{b}_{ij} \leftarrow \frac{1}{|\mathcal{V}'_k|}\sum_{(x,y)\in\mathcal{V}'_k} \ell(y, f(x \mid \mathcal{T}'_k))$ ▷ ℓ is any continuous loss function
        $\bar{b}_{ij} \leftarrow \bar{b}_{ij} + \frac{\hat{b}_{ij}}{t}$
      end for
    end for
  end for
  $\bar{b} \leftarrow \text{flatten}(\bar{b})$
  $A_{ijkl} \leftarrow \mathbb{P}(\text{Bin}(n'_{\mathcal{T}}, p_i) = k)\mathbb{P}(\text{Bin}(n'_{\mathcal{V}}, p_j) = l) \quad \forall p_i, p_j, k \in \{0, 1, \ldots, n'_{\mathcal{T}}\}, l \in \{0, 1, \ldots, n'_{\mathcal{V}}\}$
  $A \leftarrow reshape(A \in \mathbb{R}^{|\{p_i\}||\{p_j\}|\times(n'_{\mathcal{T}}+1)(n'_{\mathcal{V}}+1)})$ ▷ Each row of A is a joint binomial pmf
  $\hat{e}, residual \leftarrow A(A^\intercal A)^{-1}A^\intercal\bar{b}$
  return $\hat{e}_0, residual$
end procedure

Algorithm 4: B3 — Bidirectional leakage with unknown probabilities

procedure UNKNOWNBIDIRECTIONAL($f, \hat{\mathcal{T}}, \hat{\mathcal{V}}, n'_{\mathcal{T}}, n'_{\mathcal{V}}, t$)
  $residual^* \leftarrow \infty$
  $n_{\mathcal{T}} \leftarrow |\hat{\mathcal{T}}|, n_{\mathcal{V}} \leftarrow |\hat{\mathcal{V}}|$
  for $p_{\mathcal{T},0}$ in $\left\{\frac{0}{n_{\mathcal{T}}}, \frac{1}{n_{\mathcal{T}}}, \ldots, \frac{n_{\mathcal{T}}-1}{n_{\mathcal{T}}}\right\}$ **do**
    for $p_{\mathcal{V},0}$ in $\left\{\frac{0}{n_{\mathcal{V}}}, \frac{1}{n_{\mathcal{V}}}, \ldots, \frac{n_{\mathcal{V}}-1}{n_{\mathcal{V}}}\right\}$ **do**
      $e, residual \leftarrow$ KNOWNBIDIRECTIONAL$(f, \hat{\mathcal{T}}, \hat{\mathcal{V}}, p_{\mathcal{T},0}, p_{\mathcal{V},0}, n'_{\mathcal{T}}, n'_{\mathcal{V}}, t)$
      if $residual < residual^*$ **then**
        $e^* \leftarrow x$
        $p^*_{\mathcal{T},0} \leftarrow p_{\mathcal{T},0}$
        $p^*_{\mathcal{V},0} \leftarrow p_{\mathcal{V},0}$
        $residual^* \leftarrow residual$
      end if
    end for
  end for
  return $e^*, p^*_{\mathcal{T},0}, p^*_{\mathcal{V},0}$
end procedure